Amendments to the Claims:

Please cancel Claims 17 and 18 in favor of Claim 10 as now amended. Please amend Claims 1, 19 and 20 and please add new Claims 21-30 as shown in the below claim listing. The Claim Listing below will replace all prior versions of the claims in the application:

Claim Listing:

1.   (Currently amended) A method for analyzing a subject genome sequence comprising the steps of:

~~providing~~ <u>accessing</u> a set of known biological fragments, the set being of a fixed number of said known biological fragments, each known biological fragment in the set having a respective representation;

comparing the respective representation of each known biological fragment from the set to a subject genome sequence, for each known biological fragment said comparing including (i) counting the number of times the respective representation of the known biological fragment is found in the subject genome sequence and (ii) from said counted number of times, forming a vector element, such that for each known biological fragment there is a respective vector element representing the number of times the respective representation of that known biological fragment is found in the subject genome sequence;

from the formed vector elements, forming a vector having a length equal to the fixed number of known biological fragments in the provided set, such that the formed vector provides a uniform representation of the subject genome sequence; and

providing the formed vector for use as input to a desired analysis, the uniform representation provided by the formed vector enabling the formed vector to serve as normalized input<u>; and</u>

<u>using the uniform representation in the desired analysis, analyzing the subject genome sequence including one of classifying, clustering or indexing the subject genome sequence.</u>

2.   (Original) A method as claimed in Claim 1 wherein the set of known biological fragments is from published databases of motifs or proteins.

3.   (Previously amended) A method as claimed in Claim 1 further comprising the step of:

for each desired subject genome sequence, using said set of known biological fragments, repeating the comparing and forming steps such that a respective vector representation is formed and each desired subject genome sequence has a respective vector representation of a same length, said set of known biological fragments being a same set used for all of said subject genome sequences.

4.   (Original) A method as claimed in Claim 3 wherein for each subject genome sequence, having formed respective vector representations each of the same length, using the same length vector representation as input into one or more sequence analyses.

5.   (Original) A method as claimed in Claim 4 wherein the sequence analyses include one of indexing, classification and clustering.

6.   (Canceled)

7.   (Original) A method as claimed in Claim 1 wherein the subject genome sequence is a DNA sequence or subsequence.

8.   (Original) A method as claimed in Claim 1 wherein the counting includes determining probability of the subject genome sequence being generated by the known biological fragment.

9.   (Original) A method as claimed in Claim 8 wherein the counting determining probability employs a 0th order Markov model for each known biological fragment.

10. (Currently amended) Apparatus for analyzing genome sequences, comprising:

a data store of <u>representations of</u> a predefined number of known biological sequences;

a comparison routine executed by a digital processor having access to the data store, the comparison routine comparing each known biological sequence from the data store to a subject genome sequence and generating a score indicative of the comparison, said scores forming a vector having a length equal to the predefined number of known biological sequences, such that said comparison routine provides the formed vector as a uniform representation of the subject genome sequence and the formed vector enables at least one analysis of the subject genome sequence, the uniform representation of the formed vector providing normalized input for the analysis, <u>wherein the generated score is one of a probability of the subject genome sequence being generated by the known biological sequence or a counting of a number of occurrences of the known biological sequence found in the subject genome sequence.</u>

11. (Original) Apparatus as claimed in Claim 10 wherein the data store is a published database of motifs or proteins.

12. (Previously presented) Apparatus as claimed in Claim 10 further comprising a plurality of different subject genome sequences; and

wherein, using a same set of known biological sequences, the comparison routine forms, for each subject genome sequence, a respective vector such that a corresponding plurality of uniform length vector representations is provided.

13. (Previously presented) Apparatus as claimed in Claim 12 wherein the output of the comparison routine feeds the corresponding plurality of uniform length vector representations into further analysis processors.

14. (Original) Apparatus as claimed in Claim 13 wherein the further analysis processors include at least one of a classifier, an indexer and a clustering member.

15.   (Canceled)

16.   (Original) Apparatus as claimed in Claim 10 wherein the subject genome sequence is a DNA sequence or subsequence.

Claims 17-18 (canceled)

19.   (Currently Amended) A method for analyzing a subject protein sequence comprising the steps of:

(a)   providing a set of known biological fragments, the set being of a fixed number of said known biological fragments, each known biological fragment in the set having a respective representation;

(b)   comparing the respective representation of each known biological fragment from the set to a subject protein sequence, for each known biological fragment said comparing including (i) counting the number of times the known biological fragment is found in the subject protein sequence and (ii) from said counted number of times, forming a vector element, such that for each known biological fragment there is a respective vector element representing the number of times the respective representation of that known biological fragment is found in the subject protein sequence;

(c)   from the formed vector elements, forming a vector having a length equal to the fixed number of known biological fragments in the provided set, such that the formed vector provides a uniform representation of the subject protein sequence; and

(d)   providing the vector for making at least one analysis of the subject protein sequence, the uniform representation of the vector providing normalized input for the analysis; and

(e)   analyzing the subject protein sequence by the analysis using the uniform representation as input and classifying the subject protein sequence into a structural or a functional class.

20. (Currently Amended) Apparatus for analyzing protein sequences, comprising:

a data store of a predefined number of known biological sequences; and

a comparison routine executed by a digital processor having access to the data store, the comparison routine comparing each known biological sequence from the data store to a subject protein sequence and generating a score indicative of the comparison, said scores forming a vector having a length equal to the predefined number of known biological sequences, such that said comparison routine provides the formed vector as a uniform representation of the subject protein sequence for at least one analysis, the uniform representation of the formed vector providing normalized input for the analysis,

wherein the generated score is one of a probability of the subject protein sequence being generated by the known biological sequence or a counting of a number of occurrences of the known biological sequence found in the subject protein sequence.

21. (New) A method of analyzing a subject genome sequence comprising:

(a) providing a predefined set of known biological fragments, the set being of a fixed number of said known biological fragments, each known biological fragment in the set having a respective representation;

(b) providing a subject genome sequence;

(c) quantitatively determining a score of each known biological fragment in the set with respect to the subject genome sequence;

(d) forming a feature vector of the subject genome sequence, said feature vector being a sequence of scores of each known biological fragment in the set;

(e) using the feature vector, analyzing the subject genome sequence thereby producing classification, clustering or indexing of the subject genome sequence.

22. (New) The method of Claim 21 wherein the respective representation of each known biological fragment is a text string.

23. (New) The method of Claim 22 wherein quantitatively determining a score of each known biological fragment in the set includes for each known biological fragment, counting the number of times the text string of the respective representation is found within the subject genome sequence.

24. (New) The method of Claim 21 wherein the respective representation of each known biological fragment is a probabilistic template, said template providing a probability that a member of a group consisting of amino acids and nucleotides exists at a pre-determined position of said known biological fragment.

25. (New) The method of Claim 24 wherein quantitatively determining a score of each known biological fragment in the set includes for each known biological fragment, computing the probability of existence of every subsequence of a pre-determined length in the subject genome sequence according to the probabilistic template that represents the known biological fragment.

26. (New) Apparatus for analyzing a subject genome sequence, comprising:
    (a) an input device for inputting a subject genome sequence;
    (b) a data store of representations of a set of a predefined number of known biological fragments; and
    (c) a scoring routine executed by a digital processor having access to the data store, the scoring routine quantitatively determining a score of each known biological fragment in the set as compared against the subject genome sequence, said scores forming a feature vector having a length equal to the predefined number of known biological sequences; and
    (d) an analyzing routine executed by a digital processor, the analyzing routine using the feature vector to produce classification, indexing or clustering of the subject genome sequence.

27.   (New) The apparatus of Claim 26 wherein each known biological fragment in the set is represented by a respective text string.

28.   (New) The apparatus of Claim 27 wherein the scoring routine includes for each known biological fragment, counting the number of times the respective text string is found within the subject genome sequence.

29.   (New) The apparatus of Claim 26 wherein each known biological fragment in the set is represented by a probabilistic template, said template providing a probability that a member of a group consisting of amino acids and nucleotides exists at a pre-determined position of said known biological fragment.

30.   (New) The apparatus of Claim 29 wherein the scoring routine includes for each known biological fragment, computing the probability of existence of every subsequence of a pre-determined length in the subject genome sequence according to the probabilistic template that represents the known biological fragment.